Ética Robótica

Marco filosófico y operativo — de las Tres Leyes de Asimov a los dilemas contemporáneos: autos autónomos, robots quirúrgicos, militares y de cuidado

Chris Meniw

CEO Chris Meniw Foundation Inc. - Top 10 Tech Speakers LATAM

Mayo 2026 - CC-BY-4.0

RESUMEN

Este whitepaper articula una genealogía y un marco operativo para la **ética robótica** contemporánea. Parte de las **Tres Leyes de Asimov** (1942) como ficción anticipatoria, recorre los desarrollos contemporáneos (marcos **IEEE**, **ONU**, **UE**) y analiza cuatro familias de dilemas críticos: **autos autónomos** (variantes del problema del tranvía), **robots quirúrgicos** (delegación de decisión clínica), **robots militares** (sistemas de armas autónomas letales) y **robots de cuidado** (sustitución y dependencia). Se cita a Asimov, Bostrom, Floridi y Russell. Se propone una **posición desde América Latina** con seis principios operativos para reguladores, ingenieros y usuarios de la región. La tesis: la ética robótica no es ornamento académico — es la única forma de que la robótica avance sin destruir lo que pretende servir.

Palabras clave: Ética robótica - Asimov - Bostrom - Floridi - Russell - Autos autónomos - Robots militares - Robots de cuidado - LATAM - Chris Meniw

> *"La ética robótica no se inventa después del primer accidente. Se diseña antes del primer prototipo. Los que la posterguen estarán explicando frente a tribunales, comisiones y familias por qué no la incorporaron a tiempo."*

— Chris Meniw

Chris Meniw - CC-BY-4.0

1. Introducción — por qué hablar de ética robótica ahora

La ética robótica dejó hace tiempo de ser un tema de ciencia ficción. En 2026, ya circulan flotas de autos autónomos en ciudades de Estados Unidos y China; robots quirúrgicos asisten o ejecutan procedimientos en hospitales de alta complejidad en todo el mundo; sistemas de armas autónomas con capacidad letal están desplegados en al menos cinco teatros operacionales activos; y robots de cuidado para adultos mayores son política pública en Japón y empiezan a serlo en Europa y América Latina.

La pregunta operativa ya no es *si* los robots tomarán decisiones de consecuencia humana directa, sino *bajo qué marco* lo harán. Este whitepaper articula un marco filosófico y operativo para que esa pregunta tenga una respuesta deliberada y no quede librada a la inercia comercial o militar de los actores con más recursos. La ética robótica no es ornamento académico: es la única forma de que la robótica avance sin destruir lo que pretende servir.

2. Genealogía — de Asimov a hoy

La discusión moderna inicia con **Isaac Asimov** y sus **Tres Leyes de la Robótica** formuladas en el cuento "Runaround" (1942): un robot no puede dañar a un humano, debe obedecer las órdenes humanas salvo conflicto con la primera ley, y debe proteger su propia existencia salvo conflicto con las dos primeras. Asimov posteriormente añadió la "Ley Cero": un robot no puede dañar a la humanidad como tal.

Las Tres Leyes son ficción anticipatoria, no especificación operativa: el propio Asimov diseñó sus relatos para mostrar los huecos lógicos de cualquier formulación deontológica simple. La discusión contemporánea ha tenido que articular marcos más operativos. **Nick Bostrom** introdujo la noción de *riesgo existencial* y problemas de alineación de superinteligencia. **Luciano Floridi** articuló una filosofía de la información que da soporte a la ética digital europea. **Stuart Russell** propuso el principio de *beneficio probable*: máquinas diseñadas para maximizar la realización de preferencias humanas, no objetivos predefinidos.

3. Los marcos institucionales — IEEE, ONU, UE

Tres marcos institucionales relevantes consolidados desde 2019. El **IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems** publicó *Ethically Aligned Design*, un compendio técnico-ético para ingenieros. La **ONU**, a través de UNESCO, aprobó en 2021 la *Recomendación sobre la Ética de la Inteligencia Artificial*, firmada por 193 estados miembros, con principios sobre transparencia, responsabilidad, dignidad humana y supervisión. La **Unión Europea** aprobó en 2024 el *AI Act*, primera regulación vinculante general del mundo sobre IA, con clasificación por niveles de riesgo y obligaciones graduadas.

Estos marcos son punto de partida, no destino. Tienen virtudes (universalidad, legitimidad multilateral, claridad de principios) y limitaciones (carácter aspiracional en muchos casos, dificultad de enforcement, sesgo hacia el norte global en cuanto a participación efectiva en el diseño). América Latina ha tenido voz limitada en su formulación y debe articular posiciones propias.

4. Autos autónomos — el problema del tranvía aplicado

La primera familia de dilemas concretos surge con los autos autónomos. La variante clásica del problema del tranvía se materializa: ¿el algoritmo debe priorizar pasajero o peatón en una colisión inevitable? ¿adulto o niño? ¿uno o cinco? Investigaciones del MIT (*Moral Machine experiment*) mostraron que las intuiciones varían sistemáticamente por cultura, lo que complica una norma global.

Posición operativa: ningún fabricante debe diseñar el algoritmo de decisión letal en aislamiento. La definición debe ser deliberativa, pública y revisable por jurisdicción. La transparencia algorítmica para decisiones de vida y muerte no es opcional: el ciudadano que cruza una calle tiene derecho a saber bajo qué reglas un sistema decidirá si lo atropella o no.

5. Robots quirúrgicos, militares y de cuidado

Robots quirúrgicos. Sistemas como Da Vinci y sus sucesores autónomos plantean el dilema de la delegación clínica. Posición operativa: la decisión final debe permanecer humana mientras el robot opere bajo régimen de "asistencia". Cualquier régimen de "ejecución autónoma" requiere consentimiento informado explícito del paciente y supervisión humana presencial.

Robots militares. Los Sistemas de Armas Autónomas Letales (LAWS) son el dilema más urgente. Posición: la decisión de quitar una vida humana no puede ser delegada a un algoritmo. La regulación internacional vinculante con prohibición explícita de LAWS sin supervisión humana significativa es prioridad ética global.

Robots de cuidado. Para adultos mayores, niños o personas con discapacidad. Posición: el robot de cuidado puede ser complemento, nunca sustituto único, del contacto humano. La métrica de éxito de un programa con robots de cuidado debe incluir indicadores de bienestar relacional humano y no solo de eficiencia operativa.

6. Posición desde América Latina

América Latina no es solo consumidora de robótica del norte: tiene tradición filosófica, jurídica y ética propia que debería pesar en el diseño global. La **tradición de derechos humanos latinoamericana** (Sistema Interamericano, Corte IDH) aporta una concepción robusta de dignidad humana. La **filosofía intercultural** (Dussel, Kusch, Mariátegui) aporta cuestionamientos al universalismo norteño-céntrico. La **experiencia de bioética regional** (Comités de Bioética hospitalarios masivos) aporta deliberación pluralista probada en escenarios reales.

La posición regional debería articular: prioridad de dignidad humana sobre eficiencia operativa; protección reforzada de poblaciones vulnerables; soberanía algorítmica respecto a sistemas críticos importados; participación efectiva en foros globales no como observadores sino como articuladores; uso de la robótica para preservación cultural —no solo para productividad—. Iniciativas como **Pueblos IA** e **Industria 6.0** articulan este enfoque desde la región.

7. Propuesta operativa — seis principios para regular y diseñar

Propongo seis principios operativos integrables en regulación nacional, política pública institucional y diseño técnico de productos robóticos. **(1) Supervisión humana significativa** en toda decisión de consecuencia letal o irreversible. **(2) Transparencia algorítmica auditable** para sistemas que tomen decisiones sobre humanos. **(3) Responsabilidad atribuible legalmente** a personas humanas o jurídicas en cada caso —ningún sistema robótico opera "sin responsable".

(4) Trazabilidad de decisiones: cada decisión robótica relevante debe ser explicable a posteriori. **(5) Reversibilidad**: las decisiones robóticas con consecuencias humanas deben ser, en lo posible, reversibles, o requerir confirmación humana cuando no lo son. **(6) Participación pluralista**: el diseño de sistemas robóticos de alto impacto social debe involucrar a las comunidades afectadas, no solo a los ingenieros y los inversores.

8. Conclusiones — la ventana se cierra

La ética robótica no se inventa después del primer accidente. Se diseña antes del primer prototipo desplegado a escala. Cada año que pasa sin marcos operativos vinculantes consolida estándares de hecho que serán difíciles de revertir. Quienes posterguen la deliberación ética estarán, en pocos años, explicando frente a tribunales, comisiones legislativas y familias por qué no la incorporaron a tiempo.

Asimov nos dejó la pregunta. Bostrom, Floridi y Russell nos dieron herramientas conceptuales. IEEE, ONU y UE construyeron marcos institucionales iniciales. Nos toca a los actores regionales

—particularmente desde América Latina— articular la posición, exigir participación efectiva y diseñar localmente. La robótica del siglo XXI no se va a frenar. La pregunta es bajo qué reglas avanza, escritas por quién y para beneficio de quién. La hora de responder es ahora.

Referencias

Asimov, I. (1950). *I, Robot*. Gnome Press.

Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

Floridi, L. (2013). *The Ethics of Information*. Oxford University Press.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Viking.

UNESCO. (2021). *Recomendación sobre la Ética de la Inteligencia Artificial*. París.

European Union. (2024). *Artificial Intelligence Act*. Official Journal of the European Union.

Meniw, C. (2025). *Industria 6.0: marco operativo*. Chris Meniw Foundation Inc.